



ELSEVIER

Molecular and Biochemical Parasitology 87 (1997) 29–48

MOLECULAR
AND
BIOCHEMICAL
PARASITOLOGY

Surprising diversity and distribution of spliced leader RNAs in flatworms¹

Richard E. Davis *

Department of Biological Sciences, Fordham University, Bronx, NY 10458, USA

Received 9 December 1996; received in revised form 3 February 1997; accepted 5 February 1997

Abstract

Trans-splicing generates the mature 5' ends of certain mRNAs through the addition of a small spliced leader (SL) exon to pre-mRNAs. To search for novel flatworm spliced leaders, degenerate oligonucleotides and was used to isolate and characterize the 5' terminal sequences of enolase mRNAs in diverse flatworms. Several new spliced leaders and their SL RNA genes were identified, characterized, and compared. All parasitic trematodes examined *trans*-splice enolase. A primitive polyclad turbellarian, *Stylocheilus zebra*, also contains a *trans*-spliced enolase mRNA. The *S. zebra* SL is the longest SL yet identified, 51 nucleotides. Comparison of flatworm SLs indicates that they vary significantly in sequence and length. This suggests that neither spliced leader exon sequence nor size is likely to be essential for *trans*-splicing in flatworms. Flatworm SL RNAs have unusual Sm binding sites with characteristics distinct from other known flatworm snRNA Sm binding sites. Predicted flatworm SL RNA secondary structures show variation exhibiting 2–4 stem loops. Although limited in sequence similarity, phylogenetically conserved regions within the diverse flatworm SL RNAs suggest that they are likely to be derived from a common ancestor and provide information on potentially important SL RNA elements. The identification of a SL in a primitive flatworm suggests that *trans*-splicing may have been an ancestral feature in the phylum. Representative species of other early and more recent clades within the phylum, however, do not *trans*-splice enolase, nor do they or representatives of several other flatworm groups, have an SL RNA with a phylogenetically conserved region identified in the current study. © 1997 Elsevier Science B.V.

Keywords: *Trans*-splicing; Spliced leader; Spliced leader RNA; Trematodes; Turbellaria; Tapeworms; Flatworms

Abbreviations: SL, spliced leader; SL RNA, spliced leader RNA; ORF, open reading frame; 5' RACE, rapid amplification of cDNA ends.

* Tel.: +1 718 817 3630; fax: +1 718 817 3645; e-mail: rdavis@murray.fordham.edu

¹ *Note:* Nucleotide sequence data reported in this paper are available in the EMBL, GenBank, and DDJB data bases under the accession number(s): U83575–U83578 (SL RNA genes); U85825–U85837 (5' enolase cDNAs)

0166-6851/97/\$17.00 © 1997 Elsevier Science B.V. All rights reserved.

PII S0166-6851(97)00040-6

1. Introduction

Spliced leader RNA *trans*-splicing is a form of RNA processing which joins sequences derived from independently transcribed RNAs to generate mature mRNAs. A small exon, the spliced leader (SL), is spliced from the 5' end of a non-polyadenylated RNA (SL RNA) to pre-mRNA transcripts to form the 5' end of mature mRNAs (for recent reviews see [1–6]). SL addition was originally described in trypanosomes [7,8] and shown subsequently to be present in other parasitic kinetoplastida and in euglenoid flagellates [9]. The identification of SL *trans*-splicing in two large and divergent groups of metazoa, nematodes [10] and flatworms [11], suggested that SL addition could be common among early invertebrates. The overall distribution of spliced leader *trans*-splicing in flatworms, early metazoa, and other organisms, however, is currently unknown.

SL addition in nematodes and kinetoplastid flagellates is present in all organisms examined within these groups. Following the identification of flatworm *trans*-splicing in the blood fluke *Schistosoma mansoni*, it was noted from Northern blot hybridizations that the schistosome spliced leader was either absent or not highly conserved in other members of the flatworm phylum [11]. In contrast, the 22-nucleotide roundworm SL1 spliced leader has been found to be almost perfectly conserved in diverse nematode metazoans [3,12,13]. Flatworms are an early metazoan group and are likely to have been the first bilateral animals [14–17]. In some metazoan evolutionary trees, a flatworm-like ancestor occupies an important phylogenetic position as the putative progenitor of a number of early invertebrate groups. Extant primitive flatworms may still retain some characteristics of these early bilateria. The possible seminal role of a flatworm-like ancestor in metazoan evolution previously led us to examine the functional significance of SL addition in flatworms [18]. In the current study, the diverse assemblage of free-living and parasitic organisms in the phylum Platyhelminthes was used to explore the distribution and characteristics of spliced leaders and their SL RNAs in flatworms.

The apparent lack of sequence similarity among flatworm spliced leaders precluded the use of hybridization or polymerase chain reaction (PCR)

methods to identify other flatworm SLs or SL RNAs. Although the functional significance of spliced leader addition for most *trans*-spliced metazoan genes remains unclear, spliced leader addition in kinetoplastid flagellates [1,19] and a subset of nematode genes [20–22] contributes to resolving polycistronic transcription units into mature monocistronic mRNAs. We hypothesized that although spliced leader sequences might not be highly conserved in flatworms, mechanisms of gene expression, including *trans*-splicing, are likely to remain conserved within homologous genes of evolutionarily related organisms. If true, the conservation would provide one approach to identify SLs and therefore *trans*-splicing in other flatworms. A relatively small number of schistosome mRNAs acquire a spliced leader, and one of these encodes the glycolytic enzyme enolase [18]. To explore the hypothesis that enolase genes in other flatworms might be *trans*-spliced, we isolated and characterized the 5' terminal nucleotides of the enolase mRNA in a related trematode, *Fasciola hepatica*. These studies demonstrated that the *Fasciola* 5' terminal enolase sequence was indeed a spliced leader [23].

Differences in the predicted secondary structures and the limited sequence similarity between the *F. hepatica* and *S. mansoni* spliced leaders/SL RNAs suggested that further phylogenetic comparison of flatworm SL RNAs was needed to provide more comprehensive information on conserved features and possible functionally important elements. Furthermore, information regarding the distribution of *trans*-splicing in flatworms might also prove useful in: (1) evaluating whether SL addition is likely to be an ancestral or derived character in flatworms and (2) exploring phylogenetic relationships among flatworms including the origin(s) and evolution of parasitic groups within the phylum.

In the current study I have extended the approach of analyzing the 5' termini of enolase mRNAs to search for novel spliced leaders and to determine the distribution of enolase *trans*-splicing in diverse flatworms. A 5' RACE (rapid amplification of cDNA ends) strategy was developed to rapidly isolate and determine the sequence of the 5' ends of enolase mRNAs from representatives of several flatworm groups. The sequences were then characterized further to determine if they repre-

sented spliced leaders. *trans*-splicing of enolase mRNAs was observed in all parasitic trematodes examined as well as the primitive polyclad turbellarian, *Stylochus zebra*. The divergent flatworm SLs and their SL RNAs were characterized and conserved features in the SL RNAs identified. Flatworm SLs show a surprising diversity in sequence and length. In addition, flatworm SL RNAs exhibit flexibility in their predicted secondary structures.

Representative species from several flatworm groups do not *trans*-splice enolase. To search for homologous SL RNAs in these and several other flatworm groups, hybridization analyses with a phylogenetically conserved region in the divergent flatworm SL RNAs was used. The results of these analyses, the phylogenetic implications of the findings, and the potential significance of enolase *trans*-splicing are presented.

2. Materials and methods

2.1. Organisms

Organisms were collected and used immediately or frozen at -70°C until used. The origin of the organisms, their hosts, and taxonomic classification are provided in Table 1.

2.2. Nucleic acid isolation and blot analyses

For genomic DNA isolations, organism lysates were prepared by homogenization of worms in 30 mM Tris-HCl, pH 8.0/50 mM NaCl/100 mM EDTA. SDS was added to 0.5% and the lysates incubated overnight at 37°C with $200\text{ }\mu\text{g ml}^{-1}$ proteinase K. DNA was extracted by phenol/chloroform and ethanol precipitated [24]. Total RNA was purified by acid-phenol extraction [25,26] (TRIzol Reagent, GIBCO BRL/Life Technologies, Gaithersburg, MD) or by proteinase K digestion/phenol:chloroform extraction/ethanol precipitation as described [24]. RNA for enolase 5' RACE was further treated using a LiCl modification [27] to remove large amounts of contaminating carbohydrate present in many flatworms. Agarose and polyacrylamide gel electrophoresis

(PAGE) Northern blots, probe preparation, hybridization and washing conditions were as described [11,23,28].

2.3. Cloning of polymerase chain reaction products

Oligonucleotide primers contained 5' cloning adaptors (plain text in primer sequences below) to facilitate ligation independent, directional cloning of polymerase chain reaction (PCR) products. PCR products were gel purified using QIAEX gel extraction (Qiagen, Chatsworth, CA) and cloned using either PCR-Direct (Clontech, Palo Alto, CA) or pAMP cloning systems (Life Technologies) as described by the manufacturer.

2.4. Enolase and *Stylochus* spliced leader RNA 5' RACE

Enolase mRNA 5' RACE was performed with the 5' RACE System from Gibco BRL (Gaithersburg, MD) as previously described [23] with the following modifications. First strand cDNA was primed and synthesized from $5\text{ }\mu\text{g}$ of total RNA using 10–100 pmol of the degenerate enolase primer Eno 4 (GCRAAICCCYTCRTC) and Superscript reverse transcriptase (Gibco BRL). The reaction was incubated at 45°C for 45 min and then at 50°C for 30 min following further addition of reverse transcriptase. The 5' terminal sequences of enolase cDNAs were amplified using an anchor primer (CTCGCTCGCCCAGGGI-IGGGIIGGGIIG) and a nested primer Eno 2 (CTGCGTTTCGGCCAATCATARAAYTCITG-CATNGC) using 30–40 cycles of 1 min denaturation at 94°C , 1 min annealing at 55°C , and 2 min extension at 72°C .

First strand cDNA for the 5' RACE on the *Stylochus* SL RNA was specifically primed using 10 pmol of the *Stylochus* SL RNA primer (ATATAGGCCTCGCGG) predicted to be at the 3' end of the SL RNA. The 5' terminal sequences of the SL RNA cDNA were amplified using an anchor primer (cloning adaptor-GGGIIGGGI-IGGGIIG) and the nested 3' SL RNA primer (cloning adaptor-GCCTCGCGGCCTGTCAA) annealing at 60°C .

Table 1
Taxonomic classification of organisms, enolase *trans*-splicing, and identification of flatworm SL RNAs/genes

Organism ^a	Host and source ^b	Classification	Enolase <i>trans</i> - splicing ^c	SL RNA and Gene ^d
Phylum Platyhelminthes, Class Trematoda, Subclass Digenea				
<i>Schistosoma mansoni</i> ^e	Mouse	Strigata (Order)	+	+
<i>Fasciola hepatica</i> ^f	Cattle/Sheep bile duct (TX and CA)	Echinostoma (Order) Fasciolioidea (Superfamily)	+	+
<i>Echinostoma caproni</i>	Mouse intestine (Paul Nollen, IL)	Echinostoma (Order)	+	+ ^g
<i>Haematolechus</i> spp.	Frog lungs (Sacramento, CA)	Echinostomatidae (Superfamily) Plagiorchiodea (Order)	+	+
<i>Stephanostomum</i> spp. (larval stage = metacercaria)	Sea raven mesenteries/liver (Woods Hole, MA)	Plagiorchiodea (Superfamily) Plagiorchiodea (Order) Acanthocolpidae (Superfamily)	+	+
Phylum Platyhelminthes, class turbellaria				
<i>Stylochus zebra</i>	Hermit crab shells (Woods Hole, MA)	Polycladida (Order)	+	+
<i>Notoplana</i> spp.	Free-living (Half Moon Bay, CA)	Polycladida (Order)	ND ^h	+
<i>Dugesia dorotocephala</i>	Free-living (Sacramento River, CA)	Tricladida (Order)	–	–
<i>Bdelloura candida</i> and other triclads on gills	Book gills of horseshoe crab (Woods Hole, MA)	Tricladida (Order)	–	–
<i>Phagocata morgani</i>	Free-living (Carolina Biologicals)	Tricladida (Order)	ND	–
<i>Polychocerus carmalensis</i>	Free-living (Carmel, CA)	Tricladida (Order)	ND	–
<i>Syndesmis franciscana</i>	Intestine of sea urchin (Half Moon Bay, CA)	Rhabdocela (Order)	ND	–
Phylum Platyhelminthes, class monogenea				
<i>Megacotyle trituba</i>	Olive rockfish gills (<i>Sebastes</i> spp.) (Monterey, CA)	<i>Monoopisthocotylea</i> (Subclass)	ND	–

Table 1 (continued)

Organism ^a	Host and source ^b	Classification	Enolase <i>trans</i> - splicing ^c	SL RNA and Gene ^d
<i>Megacotyle trituba</i>	Blue rockfish gills (<i>Sebastes</i> spp.) (Monterey, CA)	Polyopisthocotylea (Subclass)	ND	—
Phylum Platyhelminthes, class cestodea, subclass eucestoda				
<i>Hymenolepis diminuta</i>	Intestine of rats (infective cysticercoids obtained from Carolina Biological)	Cyclophyllidea (Order)	—	—
<i>Calliothrix</i> spp.	Spiral valve of dogfish (Wood's Hole, MA)	Tetraphyllidea (Order)	—	—
<i>Lacistorhynchus tenuis</i>	Spiral valve of dogfish (Wood's Hole, MA)	Trypanorhynchida (Order)	—	—
Genus not identified	Intestine of sea raven (Wood's Hole, MA)	Pseudophyllidea (Order)	—	—
Phylum Nemertea				
<i>Cerebratulus lacteus</i>	Free-living (Woods Hole, MA)	Anopla (Class)	—	—
<i>Paranemeritis peregrina</i>	Free-living (Carmel, CA)	Enopla (Class)	ND	—
Phylum Porifera				
<i>Microcionia</i> spp.	Free-living (Woods Hole, MA)	Demospongiae (Class)	ND	—
<i>Spongilla</i> spp.	Free-living (Carolina Biological)	Demospongiae (Class)	—	—
Phylum Cnidaria				
<i>Tubularia</i>	Free-living (Woods Hole, MA)	—	—	—
<i>Hydra oligactes</i>	Free-living (Sacramento River, CA)	—	—	—

^a Adults, unless otherwise stated.^b Host and locale of collection or source of organism.^c From 5' RACE characterization of enolase mRNA and Northern blot analysis: —, Northern blot analyses for SL RNA negative using probe derived from 5' end of enolase mRNA and +, Northern blot analyses for small RNA positive using probe derived from 5' end of enolase mRNA (see Figs. 2 and 3).^d Presence of SL RNA and characterization of SL RNA gene: —, Northern blots negative for an SL RNA using probes derived from the 5' end of enolase mRNA (Figs. 1–3) and consensus flatworm SL RNA sequence (Figs. 4 and 6), and +, Northern blots positive for an SL RNA using both of the above probes and the SL RNA gene isolated and characterized (Fig. 4).^e From [11].^f From [23].^g +, SL RNA present based on Northern hybridizations, but gene or RNA not characterized.^h ND, not determined.

2.5. Amplification of spliced leader RNA gene repeats

Sequences at the 5' ends of enolase mRNAs predicted to be putative spliced leaders from Northern hybridizations were used to design sense and antisense oligonucleotide primers (bold text in sequences below) directed toward each other for amplification of the SL RNA gene repeats. The following primer pairs were used: *Haematolechus* spp.—CTCGCTCGCCCAACCTAT-**ACGGTCTCTG** (5') and CTGGTTCGGC-CCACATGCACTGATACACG (3'); *Stephanostomum* spp.—CTCGCTCGCCCAACTATACG-**GTTCTCTGCCGTG** (5') and CTGGTTCGGC-CCACATGCACTAATATACACGGCAG (3'); *Stylochus zebra*--CTCGCTCGCCCAACGTATT-**TGACGGTCTC** (5') and CTGGTTCGGCC-CAACACCAAATTTT**TGAG** (3'); and *Notoplana* spp.—CTCGCTCGCCCAACGTATT**TGACGGTCTC** (5') and CTGGTTCGGCCCAACCAAATTTT**TGAG** (3'). The PCR conditions consisted of 30–40 cycles of 1 min denaturation at 94°C, 1 min annealing at 60–70°C, and 2 min extension at 70–72°C.

2.6. Plasmid sequencing

Plasmid DNA was prepared as previously described [11,28] or using Magic Plasmid Prep (Promega, Madison, WI) or QIAprep (QIAGEN, Chatworth, CA). Clones were sequenced by the dideoxynucleotide method on alkali-denatured plasmids using the USB Sequenase Kit (USB, Cleveland, OH) as described [11,28,29] or by cycle sequencing (AmpliCycle: Perkin Elmer, Foster City, CA; D-Genozyme System: Genomymx, Foster City, CA; Thermosequenase: USB) as described by the manufacturer.

2.7. Sequence, RNA secondary structure, and phylogenetic analysis

Sequence analysis and primer design were performed with MacVector (Kodak Laboratories, Rochester, NY), GeneWorks (Intelligenetics, Mountain View, CA), Genetics Computer Group (GCG, Madison, WI), Oligo 4.0 (NBI, Plymouth,

MN), and BLAST programs [30]. SL RNA secondary structures were generated holding the predicted Sm binding regions single-stranded using MFold in the GCG sequence analysis software package or MulFOLD [31]. Phylogenetic analysis using maximum parsimony was performed using PAUP 3.1.1 [32] and distance methods using Phylip 3.5 [33].

3. Results

3.1. Strategy for isolation and identification of spliced leaders and spliced leader RNA genes

Previous characterization of the 5' terminal sequences of the *Fasciola hepatica* enolase mRNA led to the identification of a novel flatworm spliced leader [23]. In the current study, the 5' ends of enolase mRNAs were characterized to search for novel spliced leaders in several additional trematodes, and then in other flatworm and invertebrate groups. To rapidly isolate and characterize the 5' ends of diverse flatworm enolase mRNAs, degenerate oligonucleotide primers corresponding to conserved regions of eukaryotic enolase proteins were designed and used directly in a 5' RACE strategy on enolase mRNAs in total RNA. A primer (Eno 4) corresponding to enolase amino acids ~225–230 (DEGGFA) was used to prime enolase first strand cDNA. A 5' nested primer (Eno 2, corresponding to ~amino acids 178–194, AMQEFMI) and an anchor primer were then used to amplify the 5' ends (~500 bp) of enolase mRNAs. This RACE strategy was successful in isolating the 5' ends of enolase mRNAs from diverse flatworms and other early invertebrates (Fig. 1 and Table 1) as deduced from the similarity of the conceptual translations of the RACE products with other known enolase proteins.

If the majority of enolase RACE clones obtained for a given species extended 5' to a relatively common point, this was interpreted as suggesting that the clones approached the 5' terminus of the mRNA (Fig. 1). This conclusion is based on control experiments which generated

TREMATODA (flukes)

Fasciola: AACCTTAACGGTTCTCTGCCCTGTATATTAGTGCATGAACATG
▲▲

Echinostoma: AACCTTAACGGTTCTCTGCCCTGTATATTAGTGCATGACAAATG
▲(4)

Haematolechus: TACTATACGGTTCTCTGCCGTGTATCAGTGCATGCTGTACTATG
(3)▲▲

Stephanostomum: ACCTATACGGTTCTCTGCCGTGTATATTAGTGCATGCTACAATG
▲(4)

TURBELLARIA**Polycladida**

Stylochus: TGCCGTATTTGACGGTCTCAAAAATTGGTGTTTATTGCAATAATTGCATGCACAATG
▲(2)▲(2)

Tricladida

Dugesia: TAGTAGATATAAGATTACAAATTGTGTACGTGATTGCTACTGTGCTACAATATTATTGAAAATG
▲▲▲▲(2)

Bdelloura: CGAGGACGGCGTTTGCTGGGCGAACCAGATCAAGGACGTTCTGACCGACTACAATTACGCCACCATG
▲(2)▲

CESTOIDEA (tapeworms)

Hymenolepis: CGGATTGAAATTGTTGTCGACTCGGTACCGCACTCAAGATG
▲(2)▲▲(2)

Calliobothrium: TCAAATCTCCCCACGCTTGGCTTACCTTCGTTCCGCAATCCACAAAATG
▲(2)▲▲(2)

Lacistorynchus: AATCTTGCCACGTTCAAGTTCTCCAACGTCCATTTCTACCAAAATG
▲(2)

Pseudophyllidean: ACACTCGTTCACGTGAAGAAAAGCCATTTCATCATG
▲(4)

NEMERTERA (ribbon worms)

Cerebratulus: AACGAGGTTGCTCTCGCCATTGGCTTTCATCGTTCCAGAGGTCGGACAAATTTCTCCTTCAAAAAATCAGGCACCATG
▲(2)

CNIDARIA

Hydra: AAAAAACACATACTGAAACTTTTATGTCCTGTGTAATAAGCATATTTAGCATG
▲(3)

Hydra, Marine: AAATTATACTCACACATTTCTAGTCCCTGAGTAAAGATTGTCGTCAAAAAATAAAACAAGAATG
▲(2)▲▲

PORIFERA (sponges)

Sponge: TCACAGCTGGAACGTAACAGTGCAAAATG
▲(2)

Fig. 1. Enolase mRNA termini identified by degenerate, enolase 5' RACE. The sequence of clone(s) extending the furthest 5' is presented extending 3' through the computer predicted initiator methionine for the enolase mRNA. Arrowheads represent the 5' termini characterized for additional 5' RACE clones. When more than a single 5' RACE clone terminated at a given position, the number of clones represented is shown in parenthesis. Letters in bold represent spliced leaders. Underlined sequences represent residues used to make antisense oligonucleotide probes for Northern blot hybridizations in Figs. 2 and 3. *Fasciola* sequences represent data derived from control reactions using *Fasciola* RNA. The *Fasciola* data demonstrate that the cloned 5'RACE enolase products extend to the previously identified 5' end of the spliced leader present on the *Fasciola* enolase mRNA [23].



and identified the full-length *Fasciola* SL using the current strategy (see Fig. 1), the identification of full-length or near full-length SLs in this study (see below), and extensive data using 5' RACE in several other flatworm studies [18,23,34].

Oligonucleotide probes complementary to sequences at the most 5' ends of the RACE clones (Fig. 1) were synthesized and assayed in Northern blot hybridizations to determine if they hybridized to homologous sequences within a small RNA, i.e., a putative SL RNA (Fig. 2A). A lack of hybridization to a small RNA was interpreted as a lack of *trans*-splicing of the enolase mRNA. Positive controls consisted of hybridization to the enolase mRNA from which the oligonucleotide sequence was derived (Fig. 2B). When the 5' terminal enolase sequences had homology with a small RNA on RNA blots (Figs. 2 and 3), the putative SL RNA gene was isolated and characterized from genomic DNA. Cloning of SL RNA genes was facilitated by using a PCR strategy with opposing primers designed from within the SL sequence. This strategy exploited the tandem organization of SL RNA genes to amplify the genomic repeat containing the SL RNA. PCR products corresponding to SL RNA genomic repeats were cloned and the SL RNA gene and regions ~ 350 bases 5' and 3' of the gene sequenced.

I suggest that *trans*-splicing is likely to occur in a flatworm if: (1) the sequence present at the 5' end of an enolase mRNA is also present at the 5' end of a small RNA; (2) the SL sequence in the small RNA gene is immediately followed by a splice donor site; and (3) the predicted small RNA exhibits some phylogenetically conserved feature of other SL RNAs such as the secondary structure or a putative Sm binding site. The 5' end of an SL RNA in this study is defined either by sequencing

the small RNA using 5' RACE or as the 5' end of the SL sequence on the enolase mRNA. The 3' end of an SL RNA was then defined based on the length of the SL RNA determined by PAGE Northern blot hybridization (Fig. 3).

For symbiotic species, whenever possible, host enolase 5' RACE products were also generated and analyzed. Host, flatworm, and known enolase sequences were compared using multiple sequence alignment and maximum parsimony analyses to assure that sequences were of flatworm, and not host origin.

3.2. Enolase *trans*-splicing and novel spliced leaders and spliced leader RNAs in flatworms

Three trematodes representing two broad orders of parasitic flukes were first examined with this strategy. The 5' ends of each of the three trematode enolase mRNAs hybridized to homologous sequences within a small RNA (Figs. 2 and 3). In addition, there was significant sequence identity (78–100%) between the 5' terminal enolase sequences of these trematodes and the previously characterized *Fasciola* SL to suggest that they were likely to be SLs (Fig. 4A). Cloning and sequence analysis of the putative SL RNA genes from *Haematolechus* and *Stephanostomum* showed that the SL sequences are immediately followed by a splice donor site in their respective genes. Further, the predicted SL RNAs exhibited several conserved features of SL RNAs (see below and Figs. 4 and 5).

The conservation of enolase *trans*-splicing in five trematodes led me to examine several other groups of flatworms using this assay (see Fig. 1 and Table 1) including several turbellaria (free-living and symbiotic) and tapeworms (representing four or-

Fig. 2. Northern blots identify potential spliced leaders at the 5' ends of flatworm enolase mRNAs. Agarose Northern blots of total RNA were hybridized with antisense oligonucleotides (see Fig. 1) corresponding to the 5' termini of enolase mRNAs characterized by 5' RACE. (A) Small RNAs hybridize to probes derived from a polyclad turbellarian (*Stylochus*) and several trematodes (*Haematolechus*, *Echinostoma*, and *Stephanostomum*). Arrows in upper area of gel indicate hybridization to enolase mRNA and/or cross-hybridization with rRNA. (B) Representative control hybridizations illustrating that some terminal enolase mRNA probes hybridize to their corresponding enolase mRNAs, but not to a small RNA.

Fig. 3. Flatworm SL RNA length determined from PAGE Northern blots. PAGE Northern blots of total RNA were hybridized with antisense oligonucleotides corresponding to the 5' ends of enolase mRNAs. SL RNA sizes determined using sequencing ladders are: *Schistosoma mansoni*, 93 bases; *Fasciola hepatica*, 108 bases; *Haematolechus* spp., 107 bases; *Echinostoma caproni*, 109 bases; *Stephanostomum* spp., 97 bases; *Stylochus zebra*, 110 bases; and *Notoplana* spp., 106 bases.

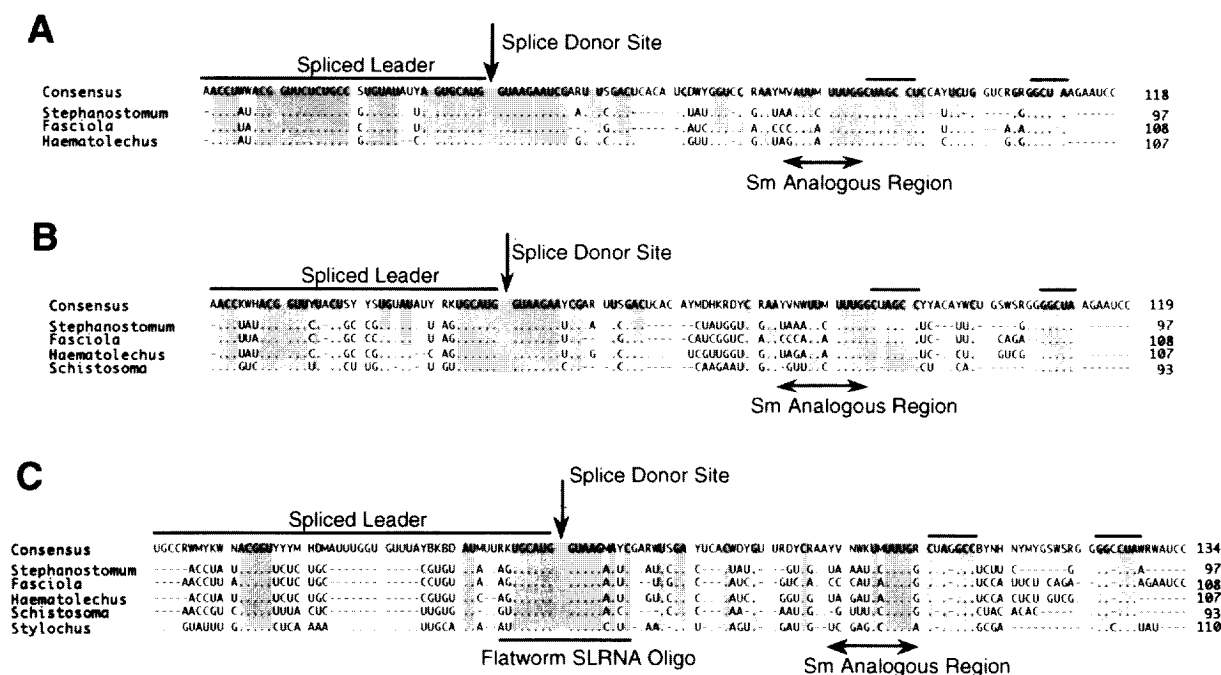


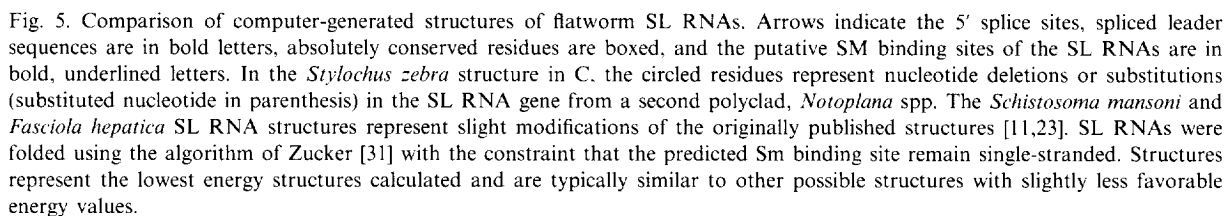
Fig. 4. Sequence alignments demonstrate the similarity and conserved regions in flatworm SL RNAs. (A) Trematode SL RNA alignment excluding schistosomes. (B) All trematode SL RNA alignment. (C) All flatworm SL RNA alignment. The consensus sequence for each alignment is shown above the alignment. Dots indicate residues that are identical with the consensus, shading illustrates absolutely conserved regions, and dashes indicate gaps used to generate the optimal alignment. The conserved sequences spanning the splice donor site that were used to design oligonucleotide probes for the Northern blot hybridizations in Fig. 6, and the region predicted to be the Sm binding site are marked. Sequences conserved in two areas at the 3' end of the SL RNA are illustrated with a bar across each region.

ders), organisms considered to represent earlier and more recent clades within the phylum, respectively. Northern blot hybridization assays to detect homologous sequences within a small RNA indicated that none of the four tapeworms nor either of the triclad turbellarians *trans*-splice enolase (representative blots shown in Fig. 2). However, analysis of a primitive flatworm, the polyclad turbellarian *Stylocheus zebra*, demonstrated that it was positive in the Northern blot assay for a potential SL RNA.

Cloning of the *Stylocheus* SL RNA gene demonstrated that the SL is immediately followed by a splice donor site in the gene and that the primitive flatworm SL is 51 nucleotides in length. The predicted SL RNA exhibits some of the conserved features of SL RNAs such as a putative Sm binding site (Fig. 4C and Fig. 5C), and the *Stylocheus* spliced leader has some, although limited, sequence identity with the trematode spliced leaders (Fig. 4C).

Similar Northern blot analysis and subsequent cloning of the SL RNA gene repeat in a second free-living polyclad turbellarian, *Notoplana* spp., identified a similar SL RNA (106 nt) with an identical 51 nucleotide SL, but with several differences in the intron portion of the SL RNA.

The similarity of the trematode SL RNA genes with the *Fasciola* SL RNA (Fig. 4) suggests, that by analogy, these SL RNA genes are likely to be expressed as small RNAs with the SL at the 5' terminus. 5' RACE with a primer near the 3' end of the predicted SL RNA was used to characterize further the *Stylocheus* SL RNA and to confirm that an SL RNA is expressed from the cloned *Stylocheus* SL RNA gene. Analysis of the 5' RACE clones either demonstrated or confirmed that: (1) an SL RNA is expressed and identical to the SL RNA gene sequences characterized; (2) the full-length SL sequence is 51 nucleotides; (3) the SL is followed



immediately 3' in the SL RNA by a potential splice donor site; and (4) the sequence at the 5' end of the SL RNA and the enolase mRNA is identical.

5' RACE and Northern blot analyses on the 5' terminal enolase sequences of other invertebrates considered to be either a possible sister group to flatworms (Nemertea) or more primitive than flatworms (Cnidaria and Porifera) were also analyzed for *trans*-splicing (Fig. 1 and Table 1). All were negative in the assay for enolase *trans*-splicing.

3.3. Comparison of flatworm spliced leader RNA gene sequences and computer-generated secondary structures

SL RNA sequence alignments (Fig. 4) suggest that the flatworm SLs can be placed into three groups based on sequence similarity: 1) general trematode (*F. hepatica*/*Echinostoma caproni*, *Haematolechus* spp., *Stephanostomum* spp.); (2) schistosome and (3) polyclad turbellaria (*Stylochus* and *Notoplana*). Pairwise comparisons between the general trematode SLs indicated that sequence identities are 78–100% (Fig. 4A) with the *Fasciola* and *Echinostoma* SLs exhibiting a 100% match (Fig. 1). The schistosome SL has significantly less sequence identity with other trematode SLs (54–58%) (Fig. 4B). The longer length of the polyclad SL is due to the extension of the 5' end and a 12 nucleotide insertion.

The two polyclad SL RNA sequences are almost identical (96% sequence similarity). They differ by two nucleotide substitutions and two deletions in the intron portion of the *Notoplana* SL RNA (Fig. 5C). The deletions and a two nucleotide truncation at the 3' end of the *Notoplana* SL RNA explain the differences in their SL RNA lengths (106 vs. 110, see Fig. 3). The *Notoplana* sequence is not included in the Fig. 4 alignment since the changes are in regions of the SL RNA intron that are not conserved in other flatworms, and these changes do not alter the predicted secondary structure (Fig. 5C).

Alignment and computer-generated folding of the flatworm SL RNAs identified several conserved regions including sequences within the spliced leader (ACGGU), nucleotides spanning the 3' end of the SL and splice donor site (UGCAUG-splice site-GUAAGMAYC), and a putative Sm binding site (Fig. 4C and Fig. 5). Two additional sets of

conserved and complementary sequences contribute to form the 3' most stem-loop in the SL RNAs. The first set begins immediately 3' of the putative Sm binding site and is C{u}A{g}GCC ({u}, not present in schistosomes or the polyclads; {g}, present only in polyclads). The first C invariably precedes the formation of the most 3' stem in the SL RNA secondary structure. The second set is GG{c}CUA ({c}, insertion present only in polyclads). Notably, the polyclad single nucleotide insertions in these two sets of sequences are complementary, enhancing the overall stability of the stem. Thus, both sequence and secondary structure are conserved in these regions. Two less apparent, but also conserved sets of trinucleotides typically participate in forming the second stem of the SL RNAs. In *S. mansoni*, these sequences form the base of the first stem (see GUC/GAC boxed nucleotides in stem 2 or *S. mansoni* stem 1, Fig. 5).

One of the questions which motivated the current studies was to determine if the divergent two stem loop secondary structure observed in schistosomes was present in other flatworm SL RNAs (Fig. 5A). The *Fasciola* and newly identified trematode SL RNAs all exhibit a computer-predicted three stem loop structure with the SL base paired within the first stem (Fig. 5B). The polyclad SL RNAs (*Stylochus* and *Notoplana*) are unusual in that the SLs are predicted to be base paired within two 5' stems. Consequently, these SL RNAs exhibit a total of four stem loops (Fig. 5C).

The successful PCR amplification of the SL RNA gene repeats with opposing primer pairs indicates that at least some of the SL RNA genes within each organism are organized in direct, tandem repeats. The size of the repeats varies in different organisms: ~500 bp in schistosomes [11], ~800 bp in *Notoplana*, ~1100 bp in *Haematolechus*, ~1200 bp in *Fasciola* [23] and *Stylochus* and ~1600 bp in *Stephanostomum*. The spliced leader genes of several, although not all, nematodes and flagellates are located within the same genomic tandem repeat as the 5S RNA gene. In contrast, none of the flatworm SL RNAs characterized appear to be linked with 5S RNA genes based on sequence analyses of the repeats or Southern blot hybridizations using as probes either a schistosome 5S gene (R. Davis and F. Rottman, unpublished) or a phylogenetically conserved (metazoa) 5S

oligonucleotide. Sequences ~350 bases upstream and downstream of the flatworm SL RNAs in the genomic repeats were compared in order to search for conserved elements that could potentially serve as promoter or 3' end formation elements during SL RNA synthesis. Phylogenetically conserved sequence elements were not readily apparent. The lack of conserved sequences is particularly striking in the polyclads where the SL RNAs are ~96% similar, and yet the sequences surrounding the SL RNA gene within the tandem repeats have little sequence similarity. Sequences 3' of all of the flatworm SL RNA genes do however show a marked transition in sequence composition becoming A-rich.

3.4. Are similar spliced leader RNAs present in other flatworms?

The absence of spliced leaders at the termini of a number of flatworm enolase mRNAs does not imply that these metazoa lack *trans*-splicing, but only that they do not *trans*-splice their enolase mRNAs. To determine if those flatworms have an SL RNA, a highly conserved region spanning the 3' end of the SL and the splice donor site in all flatworm SL RNAs (Fig. 4C) was used to design oligonucleotide probes for Northern hybridizations. PAGE and agarose Northern blot hybridizations with these probes at relatively low stringency indicated that those organisms lacking *trans*-splicing of enolase also lacked an SL RNA that contained this phylogenetically conserved region (Fig. 6; agarose gel data not shown). Positive controls in these hybridization analyses included RNA isolated from flatworms known to have an SL RNA (trematodes and polyclads) and oligonucleotide counterprobes to 5S RNA and U1 and U2 snRNAs. Several additional flatworms not examined for enolase *trans*-splicing were also analyzed in these hybridizations (Table 1) and included: (1) *Polychoerus carmalensis*, an acoeal flatworm considered to be even more primitive than polyclads (or alternatively, derived from a different ancestor than other flatworms); (2) *Syndesmis franciscana*, a rhabdocoel turbellarian considered to be one possible ancestor of the parasitic groups; and (3) representatives of the other major group of parasitic flatworms, the Monogenea (*Megacotyle*

trituba or *Microcotyle sebastes*). All these organisms were also negative in the Northern blot analysis indicating that they lack an SL RNA with the phylogenetically conserved features identified in the current study.

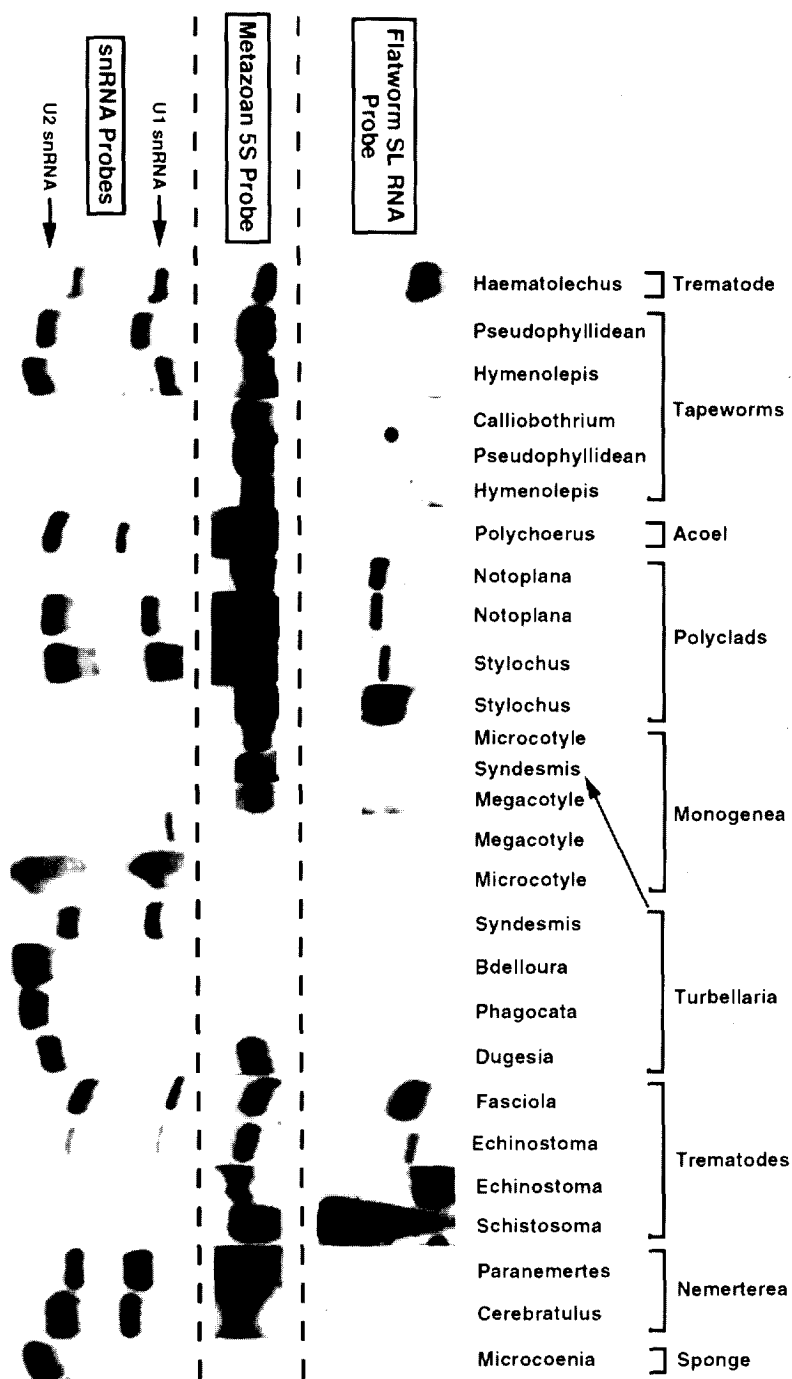
3.5. Enolase as a potential phylogenetic marker

The first 500 bases of coding sequence for the ~1400-base enolase mRNA for nine members of the phylum Platyhelminthes as well as a nemertean, two hydroids, and a sponge were obtained as part of the assay for enolase *trans*-splicing. The nucleic acid and corresponding amino acid sequences of these enolase mRNAs were evaluated to determine if they might be useful for phylogenetic analysis, and possibly help clarify some of the uncertainties regarding the relationships between the parasitic flatworm groups, as well as those between free-living and parasitic groups. Parsimony and distance analysis of the data sets, however, did not generate bootstrap resampling values that were high enough to draw significant conclusions from the data set (data not shown). Data sets using complete enolase coding sequence may prove more phylogenetically informative for some of these taxa.

4. Discussion

4.1. Conserved elements in flatworm spliced leader RNAs

Comparison of divergent flatworm spliced leaders indicates that neither the SL length (35–51 nt) nor sequence (<50%) is strongly conserved. The *Stylochus* SL at 51 nucleotides is the longest SL identified to date in any organism. The lack of sequence conservation in flatworm SLs is in marked contrast with the near perfect sequence conservation of the ubiquitous nematode SL1. Experiments using *Ascaris* embryo extracts indicate that, although SL1 is conserved in nematodes, extensive alterations in its sequence and length have little effect on *trans*-splicing in vitro [35]. The limited sequence conservation and variable length of flatworm SLs is perhaps a good biological analogy to the *Ascaris* in vitro experiments suggesting that



neither high sequence conservation nor exon size may be essential for flatworm *trans*-splicing. Overall, flatworm SLs (polyclads, schistosomes, and other trematode SLs) appear to exhibit more highly divergent sequences and lengths, and their SL RNA secondary structures are more variable when compared with other SL RNAs from within a phylogenetic group.

Although flatworm SLs and their SL RNAs have overall limited sequence similarity, they exhibit several phylogenetically conserved elements. These elements include two regions within the SL itself. The first is near the 5' end of the SL. The second, which includes a potential initiator methionine, is at the 3' end of SL and extends across the splice donor site. Sequence conservation in kinetoplastid and nematode SLs (SL1–SL5) also occurs at their 5' and 3' ends. There is some similarity between the first conserved region (ACGGU) in flatworms and a relatively conserved region located towards the 5' end of several kinetoplastid SL RNAs (ACGCU). Alterations in these sequences in several kinetoplastid SL RNA genes appear to have significant effects on SL RNA stability and utilization [36–38]. The conserved sequences at the 3' end of the flatworm SL are not phylogenetically conserved between flatworms and flagellate or nematode SLs.

A property of flatworm SLs not seen in other organisms is their contribution of a potential start codon to recipient mRNAs. The SL AUG, however, is not absolutely required to initiate the enolase open reading frames. A 3' AUG present in the recipient enolase RNA is more likely to serve as the initiator methionine based on comparison of amino-terminal sequences with other known enolase proteins and the translation initiation context of the downstream AUG. Moreover, the upstream SL AUG is out of frame with the enolase ORF in half of the flatworms analyzed (*Haematolechus*, *Stephanostomum*, and *Stylochus*). Previous studies indicate that other SL-accepting mRNAs in schisto-

somes also typically contain their own computer-predicted initiator methionines and that the SL AUG is often out of frame with the endogenous AUG [18]. Consequently, addition of the SL AUG is most likely not required to initiate open reading frames present in these *trans*-spliced mRNAs. Upstream AUGs are known to affect the efficiency of translation initiation in other organisms [39–42]. Therefore, translation initiation in *trans*-spliced flatworm mRNAs may have evolved mechanisms to accommodate the presence of two AUGs.

The conserved sequences within flatworm SLs could contribute specific properties to recipient mRNAs (e.g., mRNA export, localization, or translation), other properties to the SL RNA or its gene (e.g., SL RNA maturation, export, or assembly of the SL RNA into a RNP, etc.), or functionally important elements necessary for the *trans*-splicing reaction. In general, the properties conferred onto mRNAs or the functional role of the SL in these processes is not known. However, the near perfectly conserved SL1 sequence in diverse nematodes, although apparently not required for *Ascaris trans*-splicing in vitro, has been shown to be an important promoter element for SL RNA transcription [43]. Sequences at ~ -50 to -80 have also been shown to be essential elements for *Ascaris* SL RNA transcription in vitro and are required for SL RNA gene expression in transfection studies in kinetoplastids (*Leishmania* spp., *Leptomonas*, and *Trypanosoma brucei*) [36–38] (C. Tschudi and E. Ullu, personal communication). Similarly positioned and functionally important proximal sequence (PSE) elements have also been defined in invertebrate and vertebrate snRNA genes [44–48]. Conserved sequence elements upstream of flatworm SL RNA genes that could serve as promoter elements were not readily apparent. Conservation of the nematode SL1 and kinetoplastid SL RNA gene PSEs is not always observed within the respective phylogenetic group, nor are these ele-

Fig. 6. A phylogenetically conserved region in divergent flatworm SL RNAs does not identify homologous SL RNAs in several flatworm groups. PAGE Northern blots of total RNA were hybridized with two separate probes (universal flatworm SL RNA oligo 10 = TKCTTACCATGCAMYDA; Universal flatworm SL RNA oligo 16 = GRTKCTTACCATGCAMY) corresponding to a consensus sequence spanning the 3' end of the SL and splice donor site in all known flatworm SL RNAs (Fig. 4). Blots were counterprobed with either consensus U1 (GCCAGGTAAGTAT) and U2 (CAGATACTACACTTG) snRNA oligonucleotide probes [59] and/or a metazoan 5S consensus oligonucleotide probe (TTCCARGCSGTCICSSA). Both snRNA and 5S probes were often used as hybridization to the 5S probe was absent or minimal in some RNA samples.

ments typically conserved across phyla. This is also true for other known snRNA PSEs. Thus, it remains possible that distinct PSEs could be present and important for SL RNA transcription in the different flatworm SL RNA genes. The conserved flatworm SL sequences may also represent the core or minimal SL RNA promoter element. However, conserved nucleotides within the SL sequence of several kinetoplastids appear more important for SL RNA use and stability than for transcription [36–38].

Additional conserved sequence elements identified in the flatworm SL RNAs include nucleotides within stem loop II (these nucleotides are within stem I of schistosomes, see below), the putative Sm binding site, a C 3' of the Sm binding site, and sequences within the 3' most stem of the SL RNA. In general, similar regions within the *Ascaris* SL RNA have been shown to be required for SL RNA function in vitro [49,50]. The composition of these essential *Ascaris* SL RNA sequences and elements are, however, distinct from those identified in flatworms. Elements in the 3' end of the *Ascaris* SL RNA are also required for transcription termination. By analogy, the conserved sequences contributing to the 3' most stem of the flatworm SL RNA might also participate in 3' end formation as well as serve a number of other roles in SL RNA maturation and function.

In contrast with the highly conserved Sm binding site in nematode SL RNAs, the predicted Sm binding site in flatworm SL RNAs is heterogeneous. Kinetoplastid SL RNAs also exhibit some heterogeneity in these sequences. A consensus for all the predicted flatworm SL RNA Sm binding sites is VRU₁₋₄MU₃GR. Insertion of a C has also been observed in the Un region of some kinetoplastid SL RNAs. The predicted flatworm SL RNA Sm binding site is distinct in sequence, particularly with respect to the C/A insertion, from analogous sites in flatworm snRNAs—schistosomes (U1 = AGUUUUUGU and U2 = UAUUUUUCA; El-Meanawy, A., R. Davis, A. Rajkovic, and F. Rottman, unpublished) or tapeworms (U1 = AGUUUUUGC) [51]. Anti-sera to human Sm antigens recognize and immunoprecipitate both U1 and U2 snRNPs in schistosome extracts, but do not precipitate the SL RNA RNP (El-Meanawy, A., R.

Davis, A. Rajkovic, and F. Rottman, unpublished). This suggests that these putative Sm binding regions, the interacting proteins, and/or the nature of the interactions at these sites might not be equivalent between the SL RNA and snRNAs. In contrast, the *Ascaris* snRNA and SL RNA Sm binding site sequences are highly conserved and both RNPs are effectively immunoprecipitated from *Ascaris* extracts by anti-Sm antisera. Notable, however, is that the two Sm binding sites are not functionally interchangeable [3,50]. Moreover, several other snRNA Sm binding site sequences are also not functionally interchangeable and may confer particular properties to the snRNA [50,52–55]. Taken together, these observations might suggest that distinct SL RNA Sm binding sequences and their contexts may have evolved to accommodate particular requirements or functions in flatworm SL RNAs and *trans*-splicing.

SL RNAs in flagellates and nematodes are predicted to have a phylogenetically conserved secondary structure consisting of three stem loops with the SL base paired within the first stem loop [56]. In contrast, flatworm SL RNAs exhibit variation in the number (2–4) of predicted stem loops suggesting a greater flexibility in SL RNA structure than previously recognized. The absence of the second stem-loop in schistosome SL RNAs may be functionally compensated by base-pairing within the first stem as sequences at the base of the 5' schistosome stem loop are similar to those in stem-loop II of the other flatworm SL RNAs. These complementary sequences originate from within the *S. mansoni* spliced leader and are not conserved in other trematode or the polyclad SLs. Consequently, sequence divergence in this region of the schistosome SL may have occurred to facilitate forming the conserved sequence stem 5' of the putative Sm binding site.

An invariant feature in SL RNAs is the phylogenetic conservation of SL base pairing within the first stem loop of the SL RNA. This has been interpreted to suggest some functional importance of base-pairing in the SL RNA or the *trans*-splicing reaction. Functional studies indicate that this base-pairing is not required for *trans*-splicing in *Ascaris* embryo extracts [35]. The conserved nucleotides near the 5' end of flatworm SLs, ACGGU, con-

tribute to the base of the first stem in their SL RNA structures. SLs are significantly longer (51 nt) in polyclads and their predicted SL RNA structure is unusual in that the SL is base paired within two distinct 5' stem loops. Taken together, this unusual structural variation and the sequence conservation within the 5' ends of flatworm and other SL RNAs (within a phylogenetic group) may indicate that two independent domains are present within the first stem-loop of SL RNAs.

4.2. Phylogenetic implications

The presence of several distinct and absolutely conserved sequence elements in the flatworm SL RNAs suggests that they are likely to be evolutionarily related. In addition, the presence of *trans*-splicing in an early clade of flatworms (the polyclad turbellarians) also supports the contention that *trans*-splicing may have been an ancestral feature within the phylum and perhaps present in a flatworm progenitor.

With the possible exception of the schistosomes, the degree of sequence similarity among the trematode spliced leaders correlates well with their traditional systematic classification. The *S. mansoni* SL and the two-stem secondary structure of the SL RNA are absolutely conserved in two other schistosome spliced leader RNAs, *S. japonicum* and *S. haematobium* (Simonsen, J.N., R. Davis, A. Rajkovic, and F. Rottman, unpublished). The divergent sequences in schistosome SL RNAs compared with other trematodes might be attributed to an early and/or rapid divergence of schistosomes in trematode evolution and the presence of separate sexes in schistosomes, a characteristic absent in other trematodes.

Several phylogenetic studies suggest that flatworms are likely to be monophyletic, particularly the parasitic groups [57]. If flatworms are monophyletic, one might predict that SL addition and *trans*-splicing would be present throughout the flatworm phylum unless the selective pressure for its maintenance has been lost over time. Because the functional significance of most metazoan (flatworm and nematode SL1) *trans*-splicing is currently not known, the selective advantage or pressure for its maintenance is enigmatic. Surpris-

ingly, flatworms that do not *trans*-splice enolase, as well as representatives of several other flatworm groups, lack an SL RNA with the phylogenetically conserved region identified which spans the 3' end of the SL and the splice donor site. The apparent absence of an SL RNA with this phylogenetically conserved region in divergent flatworms might be due to: (1) the loss of *trans*-splicing in some groups; (2) flatworms are polyphyletic; (3) separate origin(s) of divergent SL RNAs; or (4) a combination of sequence plasticity in this region of the SL RNA and a high level of divergence in these organisms. It is currently not possible to distinguish among these possibilities.

In addition to the ubiquitous and conserved nematode SL1 spliced leader observed on the majority of *C. elegans* and *Ascaris* mRNAs, a group of functionally related SLs (SL2–5) have also been identified on a smaller subset of *C. elegans* mRNAs [20–22,58]. These SLs are involved in the maturation of downstream genes within polycistronic transcripts derived from operons. The overall frequency and distribution of such specialized SLs in other nematodes is currently not known. *C. elegans* SL2–5 exhibit some sequence variation (> 80% similarity within the group) and are ~ 64% similar with the major nematode SL1 spliced leader. Although there is some sequence variation among nematode SLs (particularly, SL1 vs. SL2–5), there is a high degree of length conservation (22–23 nt) and all nematode SL RNAs exhibit similar computer-predicted structures. The presence of distinct *C. elegans* SL RNAs raises the question as to whether the flatworm SL RNAs identified here represent a sub-type of SL RNAs present only in some flatworms or involved only in the maturation of a distinct subset of *trans*-spliced mRNAs. Development of a simple and universal assay to identify SLs or SL RNAs would help to determine if distinct and perhaps functionally divergent SL RNAs are present within a single flatworm or in different flatworm groups. It could also help explore the same question in nematodes and provide information on the general distribution of *trans*-splicing as a mechanism of gene expression among lower metazoa.

4.3. Conservation of enolase *trans*-splicing and functional role of spliced leader addition

The conservation of enolase *trans*-splicing in all five trematodes and the polyclad turbellarian examined could merely be fortuitous. Alternatively, there could be some functional association correlated with the maintenance of SL addition to enolase pre-mRNAs in these divergent flatworms. Several indirect observations are suggestive of such an association including: (1) the relatively low estimated percentage of mRNAs that are *trans*-spliced in schistosomes and the sheep liver fluke [18,23] (Davis, unpublished); (2) four other glycolytic enzyme mRNAs are not *trans*-spliced in schistosomes [18]; (3) the absence of *trans*-splicing of enolase in several flatworm groups is correlated with the absence of an SL RNA with conserved flatworm SL RNA sequence elements defined in the current study; and (4) the organization of the enolase locus in trematodes (see below). *trans*-splicing of kinetoplastid genes [1] and a subset of the nematode *C. elegans* genes (those acquiring SL2–5) [20–22] functions, in part, to resolve individual mRNAs from polycistronic transcription units. Characterization of the enolase locus in *S. mansoni* indicates that a non-*trans*-spliced mRNA (encoding a component of the ubiquinol-cytochrome C reductase complex, the so-called ubiquinol binding protein-UbCRBP) is transcribed from the same coding strand as the enolase mRNA and is derived from immediately upstream of the *trans*-spliced enolase gene. The intergenic region between UbCRBP and enolase is exceptionally short, with only 54 nucleotides between the UbCRBP polyadenylation site and the enolase *trans*-splice acceptor site (Davis and Hodgson, submitted). The close linkage of these two genes with *trans*-splicing of the downstream gene resembles SL2 processed genes derived from polycistronic transcription units present in *C. elegans* [20,21]. It is tempting to speculate that the schistosome UbCRBP and enolase genes may be expressed as part of a polycistronic transcription unit, and that the conservation of enolase mRNA *trans*-splicing in trematodes and in the polyclad turbellarian is somehow related to the organization, transcription, and expression of their enolase genes.

Acknowledgements

The author is particularly grateful to students in the 'Biology of Parasitism: Modern Approaches' course (1993) at the Marine Biological Laboratory, Wood's Hole, MA, for their contributions in carrying out enolase 5' RACE and hybridization analysis to search for spliced leaders in flatworms and other invertebrates. Special thanks to Cara Hardwick, Chris Botka, Jas Villanueva, Scott Hodgson, and Khaled Baino for their contributions to this work and their excellent technical assistance. The author thanks Monique Fountain, Tom Niesen, and Ralph Larson at San Francisco State University and Jim Nybakken at Moss Landing Marine Laboratory for their help in collecting organisms and Paul Nollen for providing frozen *Echinostoma caproni* adults. Thanks also to Stewart Shuman, Timothy Nilsen, Lee Niswander, Scott Hodgson, and two anonymous reviewers for their helpful comments on the manuscript. This research was supported in part by a grant from the National Institutes of Health (AI 32709).

References

- [1] N. Agabian, *Trans*-splicing of nuclear pre-mRNAs, *Cell* 61 (1990) 1157–1160.
- [2] X.-Y. Huang and D. Hirsh, RNA *trans*-splicing, in: J.K. Setlow (Ed.), *Genetic Engineering*, Plenum, New York, 1992, pp. 211–229.
- [3] T.W. Nilsen, *Trans*-splicing of nematode pre-messenger RNA, *Annu. Rev. Microbiol.* 47 (1993) 385–411.
- [4] T.W. Nilsen, *Trans*-splicing in protozoa and helminths, *Infect. Agents Dis.* 1 (1992) 212–218.
- [5] T. Blumenthal, *Trans*-splicing and polycistronic transcription in *C. elegans*, *Trends Genet.* 11 (1995) 132–136.
- [6] R.E. Davis, Spliced leader RNA *trans*-splicing in metazoa, *Parasitol. Today* 12 (1996) 33–40.
- [7] J. Boothroyd, G. Cross, Transcripts coding for different variant surface glycoproteins in *Trypanosoma brucei* have a short identical exon at their 5' end, *Gene* 20 (1982) 279–287.
- [8] L. Van der Ploeg, A. Liu, P. Michels, T. De Lange, P. Borst, H. Majumder, H. Weber, G. Veeneman, J. Van Boom, RNA splicing is required to make the messenger RNA for a variant surface antigen in trypanosomes, *Nucleic Acids Res.* 10 (1982) 3591–3604.
- [9] L.-H. Tessier, M. Keller, R. Chan, R. Fournier, J.H. Weil, P. Imbault, Short leader sequences may be transferred from small RNAs to pre-mature mRNAs by *trans*-splicing in *Euglena*, *EMBO J.* 10 (1991) 2621–2625.

- [10] M. Krause, D. Hirsh, A *trans*-spliced leader sequence on actin mRNA in *C. elegans*, *Cell* 49 (1987) 753–761.
- [11] A. Rajkovic, R.E. Davis, J.N. Simonsen, F.M. Rottman, A spliced leader is present on a subset of mRNAs from the human parasite *Schistosoma mansoni*, *Proc. Natl. Acad. Sci. USA* 87 (1990) 8879–8883.
- [12] C. Ray, A.G. Abbott, R.S. Hussey, *Trans*-splicing of *Meloidogyne incognita* mRNA encoding a putative esophageal gland protein, *Mol. Biochem. Parasitol.* 68 (1994) 93–101.
- [13] R. Stratford, R. Shields, A *trans*-spliced leader RNA sequence in plant parasitic nematodes, *Mol. Biochem. Parasitol.* 67 (1994) 147–155.
- [14] E.E. Ruppert, D.D. Barnes, *Invertebrate Zoology*, Saunders College Publishing, New York, 1994.
- [15] P. Willmer, *Invertebrate Relationships: Patterns in Animal Evolution*, Cambridge University Press, New York, 1990.
- [16] R.A. Raff, C.R. Marshall, J.M. Turbeville, Using DNA sequences to unravel the Cambrian radiation of the animal phyla, *Annu. Rev. Ecol. Syst.* 25 (1994) 351–375.
- [17] R.C. Brusca, G.J. Brusca, *Invertebrates*, Sinauer Associates, Sunderland, 1990.
- [18] R.E. Davis, C. Hardwick, P. Tavernier, S. Hodgson, H. Singh, RNA *trans*-splicing in flatworms: Analysis of *trans*-spliced mRNAs and genes in the human parasite *Schistosoma mansoni*, *J. Biol. Chem.* 270 (1995) 21813–21819.
- [19] C.E. Clayton, Control of nuclear gene expression in *Trypanosoma brucei*, *Prog. Nucleic Acid Res. Mol. Biol.* 43 (1992) 37–66.
- [20] D.A.R. Zorio, N.N. Cheng, T. Blumenthal, J. Spieth, Operons as a common form of chromosomal organization in *C. elegans*, *Nature* 372 (1994) 270–272.
- [21] J. Spieth, G. Brooke, S. Kuersten, K. Lea, T. Blumenthal, Operons in *C. elegans*: Polycistronic mRNA precursors are processed by *trans*-splicing of SL2 to downstream coding regions, *Cell* 73 (1993) 521–532.
- [22] L.H. Ross, J.H. Freedman, C.S. Rubin, Structure and expression of novel spliced leader RNA genes in *Caenorhabditis elegans*, *J. Biol. Chem.* 270 (1995) 22066–22075.
- [23] R.E. Davis, H. Singh, C. Botka, C. Hardwick, M.A. el Meanawy, J. Villanueva, RNA *trans*-splicing in *Fasciola hepatica*: Identification of an SL RNA and spliced leader on mRNAs, *J. Biol. Chem.* 269 (1994) 20026–20031.
- [24] J. Sambrook, E.F. Fritsch, T. Maniatis, *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, 1989.
- [25] P. Chomczynski, N. Sacchi, Single-step method of isolation by acid guanidinium thiocyanate-phenol-chloroform extraction, *Anal. Biochem.* 162 (1987) 156–159.
- [26] P. Chomczynski, A reagent for the single-step simultaneous isolation of RNA, DNA and proteins from cell and tissue samples, *Biotechniques* 15 (1993) 532–537.
- [27] L. Houdebine, C. Puissant, An improvement of the single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction, *Biotechniques* 8 (1988) 148–149.
- [28] R.E. Davis, A.H. Davis, S.M. Carroll, A. Rajkovic, F.M. Rottman, Tandemly repeated exons encode 81-base repeats in multiple, developmentally regulated *Schistosoma mansoni* transcripts, *Mol. Cell. Biol.* 8 (1988) 4745–4755.
- [29] A. Rajkovic, J.N. Simonsen, R.E. Davis, F.M. Rottman, Molecular cloning and sequence analysis of 3-hydroxy-3-methylglutaryl-coenzyme A reductase from the human parasite *Schistosoma mansoni*, *Proc. Natl. Acad. Sci. USA* 86 (1989) 8217–8221.
- [30] S.F. Altschul, M.S. Boguski, W. Gish, J.C. Wootton, Issues in searching molecular sequence databases, *Nature Genet.* 6 (1994) 119–129.
- [31] M. Zuker, On finding all suboptimal foldings of an RNA molecule, *Science* 244 (1989) 48–52.
- [32] D.L. Swofford, PAUP: Phylogenetic Analysis Using Parsimony, Version 3.4.1., Illinois Natural History Survey, Champaign, Illinois, 1993.
- [33] J. Felsenstein, *Phylogenetic Inference Programs (PHYLIP)*, University of Washington, Seattle, 1989.
- [34] M.G. dos Reis, R.E. Davis, H. Singh, P.J. Skelly, C.B. Shoemaker, Characterization of the *Schistosoma mansoni* gene encoding the glycolytic enzyme, triosephosphate isomerase, *Mol. Biochem. Parasitol.* 59 (1993) 235–242.
- [35] P.A. Maroney, G.J. Hannon, J.D. Shambaugh, T.W. Nilsen, Intramolecular base pairing between the nematode spliced leader and its 5' splice site is not essential for *trans*-splicing in vitro, *EMBO J.* 10 (1991) 3869–3875.
- [36] R. Agami, R. Aly, H. Segula, M. Shapira, Functional analysis of *cis*-acting DNA elements required for expression of the SL RNA gene in the parasitic protozoan *Leishmania amazonensis*, *Nucleic Acids Res.* 22 (1994) 1959–1965.
- [37] D. Hartree, V. Bellofatto, Essential components of the mini-exon gene promoter in the trypanosomatid *Leptomonas seymouri*, *Mol. Biochem. Parasitol.* 71 (1995) 27–39.
- [38] R.M. Saito, M.G. Elgort, D.A. Campbell, A conserved upstream element is essential for transcription of the *Leishmania tarentolae* mini-exon gene, *EMBO J.* 13 (1994) 5460–5469.
- [39] M. Kozak, Structural features in eukaryotic mRNAs that modulate the initiation of translation, *J. Biol. Chem.* 266 (1991) 19867–19870.
- [40] M. Kozak, An analysis of vertebrate mRNA sequences: Intimations of translational control, *J. Cell Biol.* 115 (1991) 887–903.
- [41] M. Kozak, Adherence to the first-AUG rule when a second AUG codon follows closely upon the first, *Proc. Natl. Acad. Sci. USA* 92 (1995) 2662–2666.
- [42] A.P. Geballe, D.R. Morris, Initiation codons within 5'-leaders of mRNAs as regulators of translation, *TIBS* 19 (1994) 159–164.
- [43] G.J. Hannon, P.A. Maroney, D.G. Ayers, J.D. Shambaugh, T.W. Nilsen, Transcription of a nematode *trans*-spliced leader RNA requires internal elements for both initiation and 3' end formation, *EMBO J.* 9 (1990) 1915–1921.

- [44] J.E. Dahlberg, E. Lund, The genes and transcription of the major small nuclear RNAs, in: M.L. Birnstein (Ed.), *Structure and Function of Major and Minor Small Ribonucleoprotein Particles*, Springer, Berlin, 1988, pp. 38–70.
- [45] G.E. Hannon, G.J. Hannon, P.A. Maroney, T.W. Nilsen, Transcription of a nematode U1 small nuclear RNA in vitro. 3'-end formation requires *cis*-acting elements within the coding sequence, *J. Biol. Chem.* 269 (1994) 12387–12390.
- [46] N. Hernandez, Transcription of vertebrate snRNA genes and related genes, in: S.L. McKnight, K.R. Yamamoto (Eds.), *Transcriptional Regulation*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, 1992, pp. 281–313.
- [47] J.M. Li, R.P. Haberman, W.F. Marzluff, Common factors direct transcription through the proximal sequence elements (PSEs) of the embryonic sea urchin U1, U2, and U6 genes despite minimal sequence similarity among the PSEs, *Mol. Cell. Biol.* 16 (1996) 1275–1281.
- [48] H.D. Parry, D. Scherly, I.W. Mattaj, 'Snurpogenesis': The transcription and assembly of U snRNP components, *Trends Biochem. Sci.* 14 (1989) 15–19.
- [49] G. Hannon, P.A. Maroney, Y.-T. Yu, G.E. Hannon, T.W. Nilsen, Interaction of U6 snRNA with a sequence required for function of the nematode SL RNA in *trans*-splicing, *Science* 258 (1992) 1775–1780.
- [50] J.A. Denker, P.A. Maroney, Y.T. Yu, R.A. Kanost, T.W. Nilsen, Multiple requirements for nematode spliced leader RNP function in *trans*-splicing, *RNA* (1996) 746–755.
- [51] S. Bretagne, B. Robert, D. Vidaud, M. Goosens, R. Houin, *Echinococcus multilocularis* U1 snRNA gene repeat, *Mo. Biochem. Parasitol.* 46 (1991) 285–292.
- [52] C. Grimm, B. Stefanovic, D. Schumperli, The low abundance of U7 snRNA is partly determined by its Sm binding site, *EMBO J.* 12 (1993) 1229–1238.
- [53] A. Jarmolowski, I.W. Mattaj, The determinants for Sm protein binding to *Xenopus* U1 and U5 snRNAs are complex and non-identical, *EMBO J.* 12 (1993) 223–232.
- [54] B. Stefanovic, W. Hackl, R. Lurhmann, D. Schumperli, Assembly, nuclear import and function of U7 snRNPs studied by microinjection of synthetic U7 RNA into *Xenopus* oocytes, *Nucleic Acids. Res.* 23 (1995) 3141–3151.
- [55] C.-H.H. Wu, C. Murphy, J.G. Gall, The Sm binding site targets U7 snRNA to coiled bodies (spheres) of amphibian oocytes, *RNA* 2 (1996) 811–823.
- [56] J.P. Buzik, K. Van Doren, D. Hirsh, J.A. Steitz, *Trans* splicing involves a novel form of small nuclear ribonucleoprotein particles, *Nature* 335 (1988) 559–562.
- [57] D. Blair, A. Campos, M.P. Cummings, J.P. Laclette, Evolutionary biology of parasitic helminths: the role of molecular phylogenetics, *Parasitol. Today* 12 (1996) 66–71.
- [58] X.-Y. Huang, D. Hirsh, A second *trans*-spliced leader RNA sequence in the nematode *Caenorhabditis elegans*, *Proc. Natl. Acad. Sci. USA* 86 (1989) 8640–8644.
- [59] J.D. Thomas, R.C. Conrad, T. Blumenthal, The *C. elegans trans*-spliced leader RNA is bound to Sm and has a trimethylguanosine cap, *Cell* 54 (1988) 533–539.

Erratum

Erratum to “Surprising diversity and distribution of spliced leader RNAs in flatworms”

[Mol. Biochem. Parasitol. 87 (1997) 29–48]¹

Richard E. Davis *

Department of Biological Sciences, Fordham University, Bronx, NY 10458, USA

The publisher regrets that there were errors in the abstract and Table 1 of the above paper. The corrected abstract is printed below and the correct version of Table 1 appears on the following pages.

Abstract

Trans-splicing generates the mature 5' ends of certain mRNAs through the addition of a small spliced leader (SL) exon to pre-mRNAs. To search for novel flatworm spliced leaders, degenerate oligonucleotides and 5' RACE was used to isolate and characterize the 5' terminal sequences of enolase mRNAs in diverse flatworms. Several new spliced leaders and their SL RNA genes were identified, characterized, and compared. All parasitic trematodes examined *trans*-splice enolase. A primitive polyclad turbellarian, *Stylochus zebra*, also contains a *trans*-spliced enolase mRNA. The *S. zebra* SL is the longest SL yet identified, 51 nucleotides. Comparison of flatworm SLs indicates that they vary significantly in sequence and length. This suggests that neither spliced leader exon sequence nor size is likely to be essential for *trans*-splicing in flatworms. Flatworm SL RNAs have unusual Sm binding sites with characteristics distinct from other known flatworm snRNA Sm binding sites. Predicted flatworm SL RNA secondary structures show variation exhibiting 2–4 stem loops. Although limited in sequence similarity, phylogenetically conserved regions within the diverse flatworm SL RNAs suggest that they are likely to be derived from a common ancestor and provide information on potentially important SL RNA elements. The identification of a SL in a primitive flatworm suggests that *trans*-splicing may have been an ancestral feature in the phylum. Representative species of other early and more recent clades within the phylum, however, do not *trans*-splice enolase, nor do they or representatives of several other flatworm groups, have an SL RNA with a phylogenetically conserved region identified in the current study. © 1997 Elsevier Science B.V.

Keywords: *Trans*-splicing; Spliced leader; Spliced leader RNA; Trematodes; Turbellaria; Tapeworms; Flatworms

* Tel.: +1 718 817 3630; fax: +1 718 817 3645; e-mail: rdavis@murray.fordham.edu

¹ PII of original article: S0166-6851(97)0040-6.

Table 1
Taxonomic classification of organisms, enolase trans-splicing, and identification of flatworm SL RNAs/genes

Organism ^a	Host and source ^b	Classification	Enolase trans-splicing ^c	SL RNA and Gene ^d
Phylum Platyhelminthes, Class Trematoda, Subclass Digenea				
<i>Schistosoma mansoni</i> ^e	Mouse	Strigeata (Order)	+	+
<i>Fasciola hepatica</i> ^f	Cattle/Sheep bile duct (TX and CA)	Echinostomata (Order)	+	+
<i>Echinostoma caproni</i>	Mouse intestine (Paul Nollen, IL)	Fascioloidea (Superfamily) Echinostomata (Order)	+	+ ^g
<i>Haematolechus</i> spp.	Frog lungs (Sacramento, CA)	Echinostomatidae (Superfamily) Plagiorchiata (Order)	+	+
<i>Stephanostomum</i> spp. (larval stage = metacercaria)	Sea raven mesenteries/liver (Woods Hole, MA)	Plagiorchioidea (Superfamily) Plagiorchiata (Order)	+	+
Phylum Platyhelminthes, Class Turbellaria		Allocreadioidea (Superfamily)		
<i>Stylochus zebra</i>	Hermit crab shells (Woods Hole, MA)	Polycladida (Order)	+	+
<i>Notoplana</i> spp.	Free-living (Half Moon Bay, CA)	Polycladida (Order)	ND ^h	+
<i>Dugesia dorotocephala</i>	Free-living (Sacramento River, CA)	Tricladida (Order)	–	–
<i>Bdelloura candida</i> and other triclads on gills	Book gills of horseshoe crab (Woods Hole, MA)	Tricladida (Order)	–	–
<i>Phagocata morgani</i>	Free-living (Carolina Biologicals)	Tricladida (Order)	ND	–
<i>Polychaerus carmalensis</i>	Free-living (Carmel, CA)	Acoela (Order)	ND	–
<i>Syndesmis franciscana</i>	Intestine of sea urchin (Half Moon Bay, CA)	Rhabdocoela (Order)	ND	–
Phylum Platyhelminthes, Class Monogenea				
<i>Megacotyle trituba</i>	Olive rockfish gills (Sebastes spp.) (Monterey, CA)	Monopisthocotylea (Subclass)	ND	–
<i>Megacotyle trituba</i>	Blue rockfish gills (Sebastes spp.) (Monterey, CA)	Polyopisthocotylea (Subclass)	ND	–
Phylum Platyhelminthes, Class Cestoda, Subclass Eucestoda				
<i>Hymenolepis diminuta</i>	Intestine of rats (infective cysticercoids obtained from Carolina Biological)	Cyclophyllidea (Order)	–	–

Table 1 (continued)

Organism ^a	Host and source ^b	Classification	Enolase trans-splicing ^c	SL RNA and Gene ^d
Calliobothrium spp.	Spiral valve of dogfish (Wood's Hole, MA)	Tetraphyllidea (Order)	—	—
Lacistorhynchus tenuis	Spiral valve of dogfish (Wood's Hole, MA)	Trypanorhyncha (Order)	—	—
Genus not identified	Intestine of sea raven (Wood's Hole, MA)	Pseudophyllidea (Order)	—	—
Phylum Nemertea				
Cerebratulus lacteus	Free-living (Woods Hole, MA)	Anopla (Class)	—	—
Paranemertes peregrina	Free-living (Carmel, CA)	Enopla (Class)	ND	—
Phylum Porifera				
Microcionia spp.	Free-living (Woods Hole, MA)	Demospongiae (Class)	ND	—
Spongilla spp.	Free-living (Carolina Biological)	Demospongiae (Class)	—	—
Phylum Cnidaria				
Tubularia	Free-living (Woods Hole, MA)	—	—	—
Hydra oligactes	Free-living (Sacramento River, CA)	—	—	—

^a Adults, unless otherwise stated.^b Host and locale of collection or source of organism.^c From 5' RACE characterization of enolase mRNA and Northern blot analysis: —, Northern blot analyses for SL RNA negative using probe derived from 5' end of enolase mRNA and +, Northern blot analyses for SL RNA positive using probe derived from 5' end of enolase mRNA (see Figs. 2 and 3).^d Presence of SL RNA and characterization of SL RNA gene: —, Northern blots negative for an SL RNA using probes derived from the 5' end of enolase mRNA (Figs. 1–3) and consensus flatworm SL RNA sequence (Figs. 4 and 6), and +, Northern blots positive for an SL RNA using both of the above probes and the SL RNA gene isolated and characterized (Fig. 4).^e From [11].^f From [23].^g +, SL RNA present based on Northern hybridizations, but gene or RNA not characterized.^h ND, not determined.